

Integrating SQL queries: from PROC SQL to SQLDF

Julia Psenner
Statistician
mainanalytics GmbH
julia.psenner@mainanalytics.de

Effective data processing → SQL queries

Include SQL queries in other languages:

SAS: PROC SQL

- Most common way to use SQL queries in clinical context

R: SQLDF

- Growing popularity due to shift to open source programming
- Easy way for traditional SAS users to perform SQL queries in R

Basic call to create new_table from old_table

PROC SQL

```
proc sql;
  create table new_table as
  select * from old_table;
quit;
```

- No need for package
- Create new table with statement "create table ... as"
- SQL statement in one data step

SQLDF

```
install.packages("sqldf")
library(sqldf)
new_table <- sqldf("select * from old_table")
```

- Need to install and load package "sqldf"
- Create new table with assignment ("<-" or "=")
- SQL statement in parenthesis and quotation mark

Database Management System (DBMS): when executing SQL statement → SQL database (DB) is created in backend

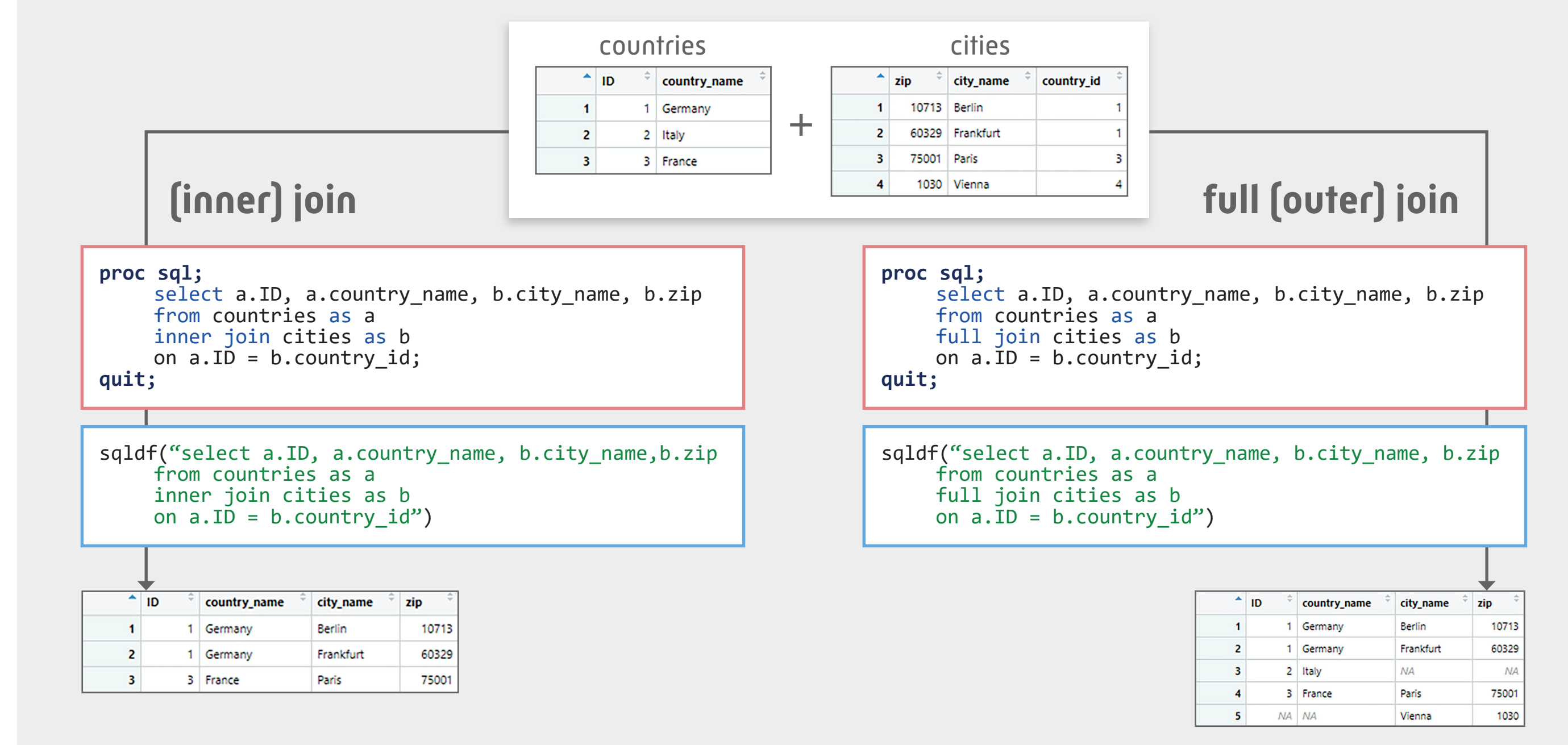
If PROC SQL and SQLDF connect to DB of same DBMS, there is no different behavior between the two languages.

This poster shows the similarities and differences when working on the default DBMS:

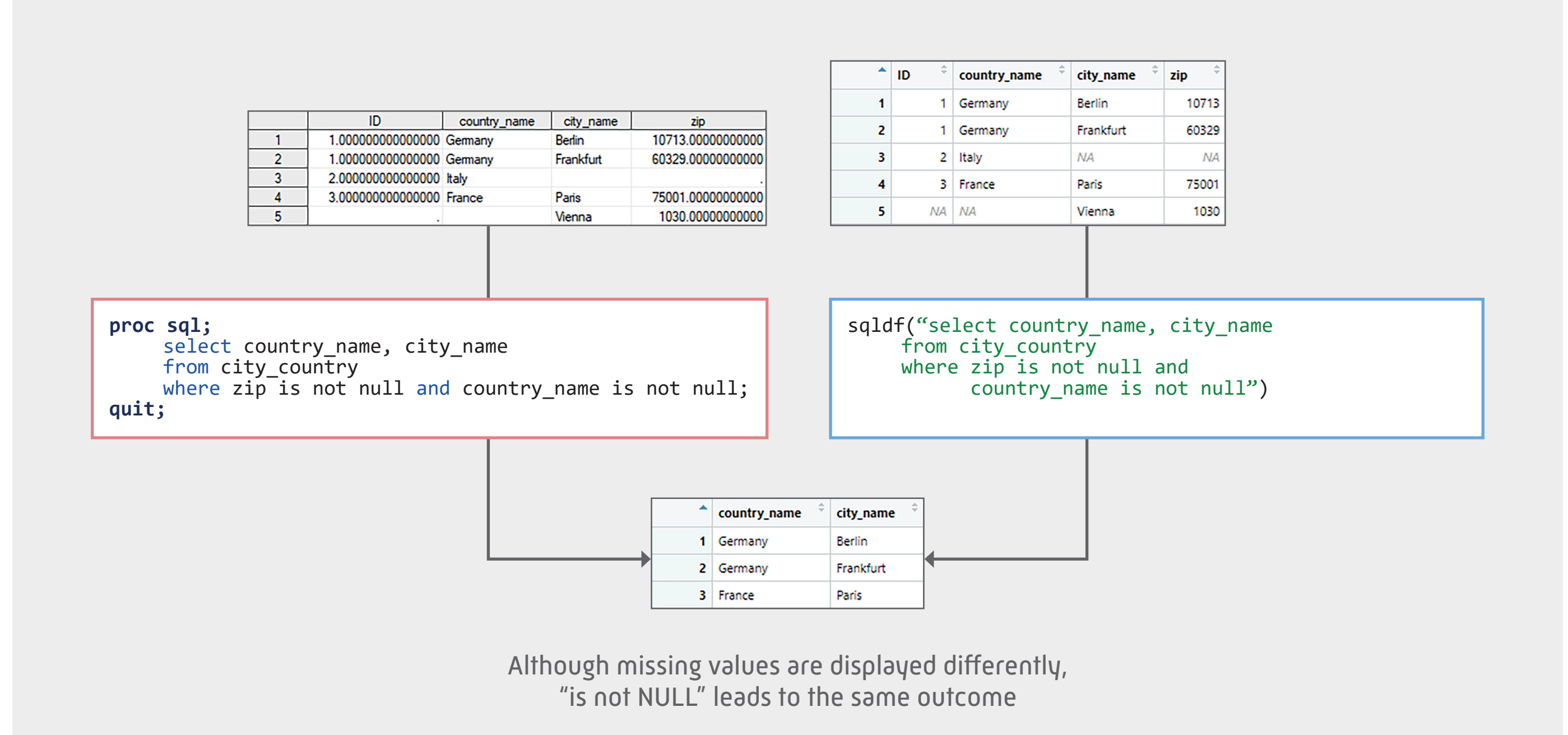
- PROC SQL:** internal database
- SQLDF:** connects to SQLite database

Similarities

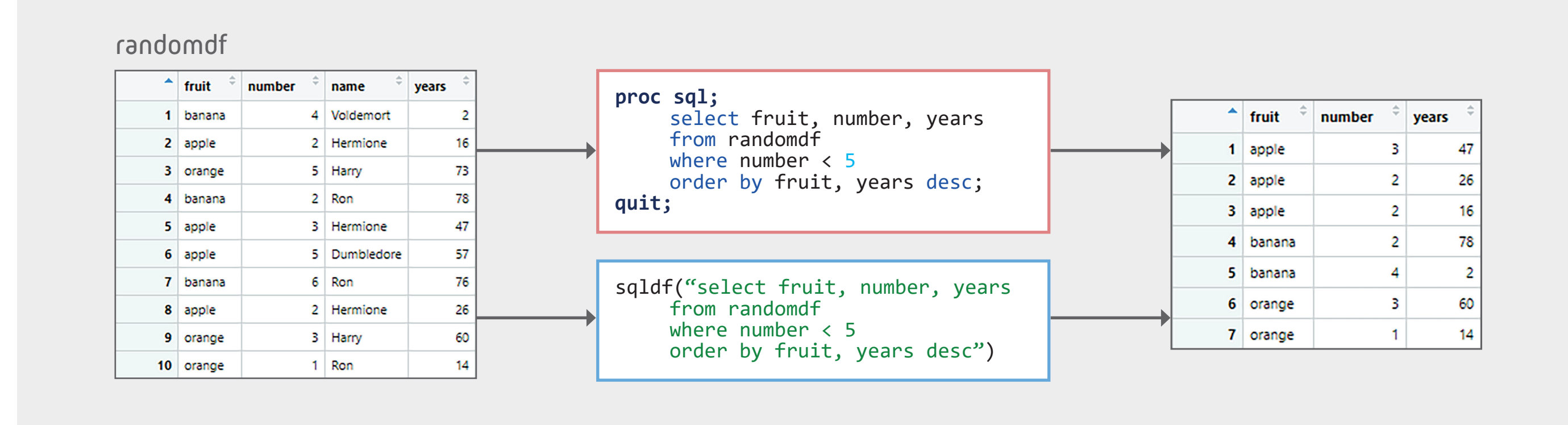
Join two datasets



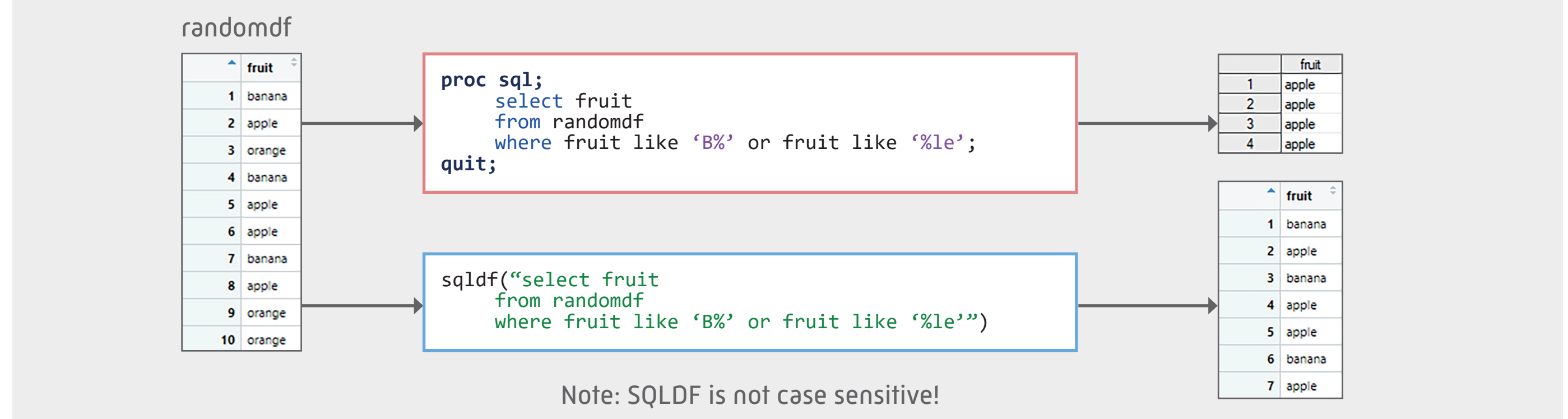
Missing values



Select, where, order by

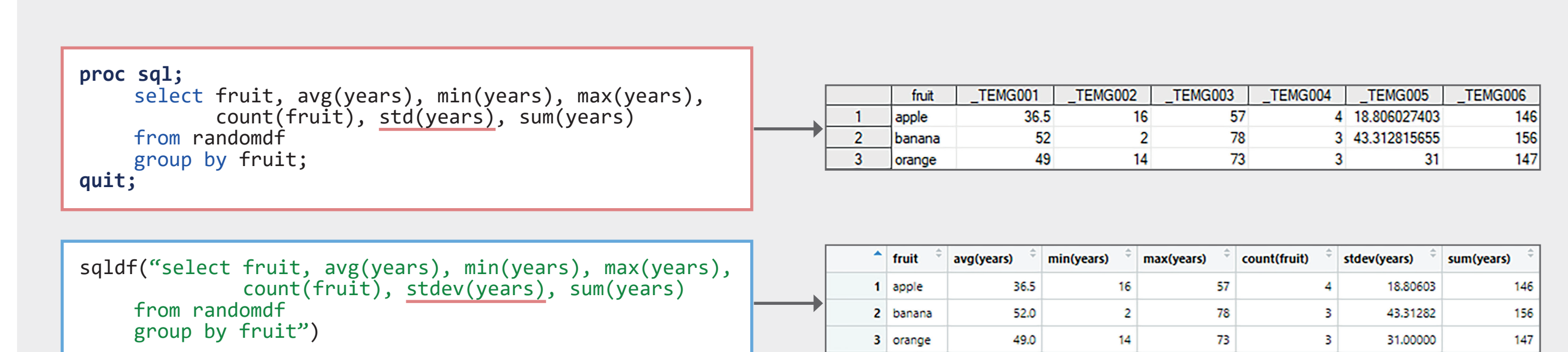


Text operators



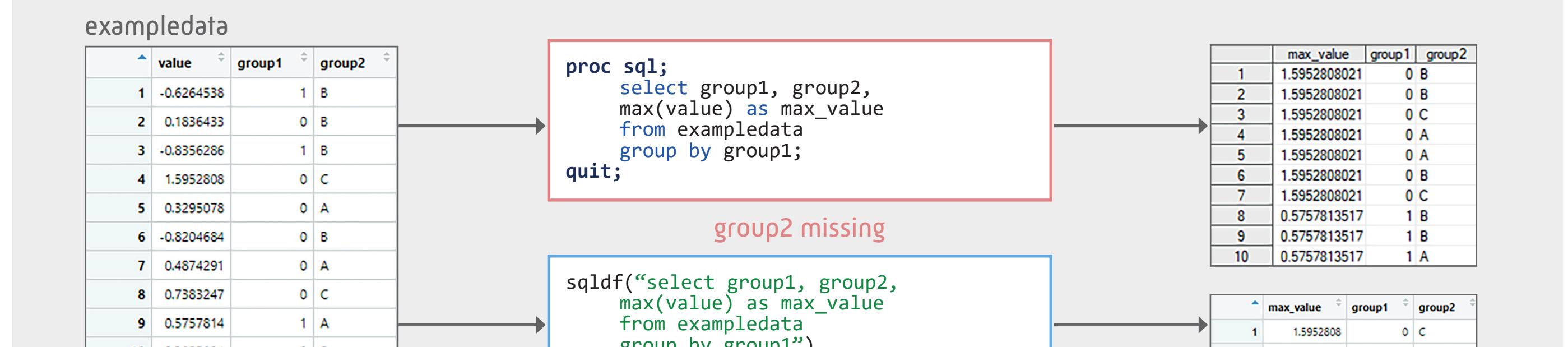
Differences

Aggregate functions



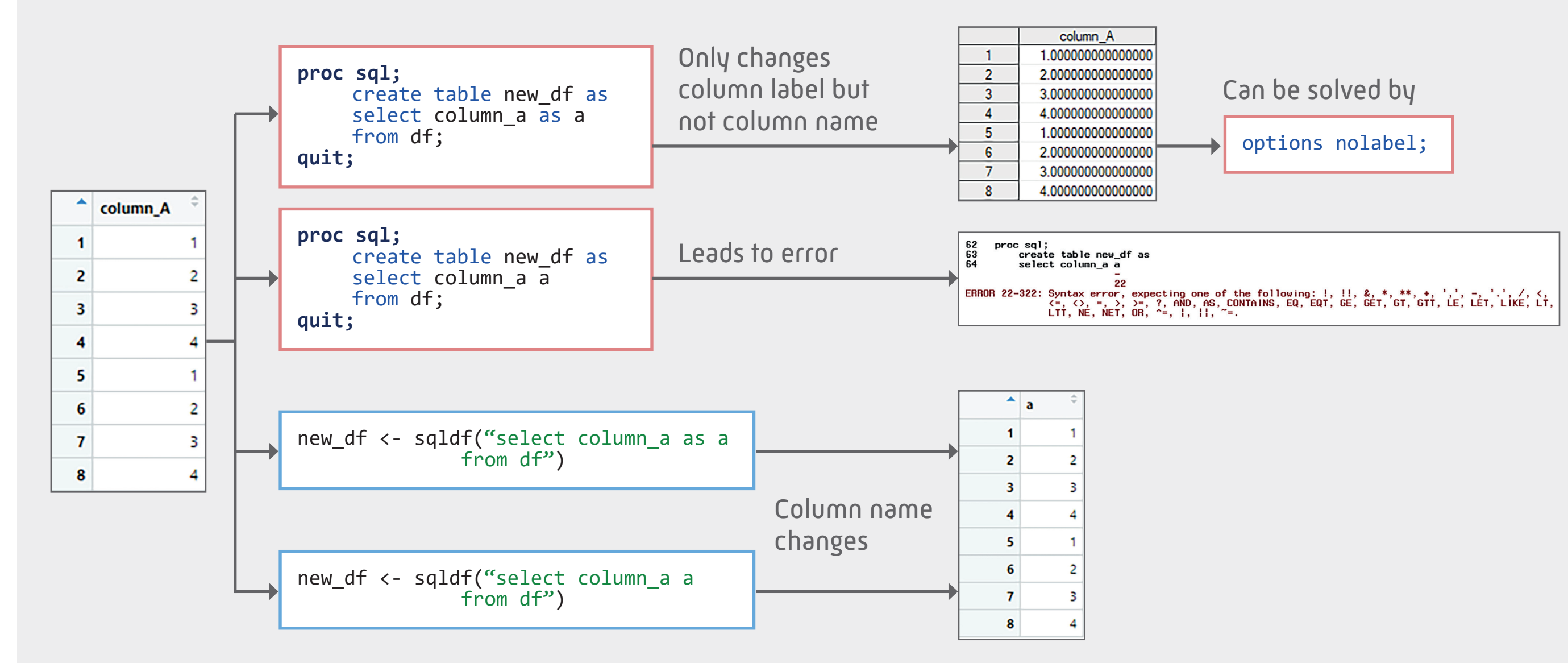
Aggregate functions work the same way, but unlike in PROC SQL, the new variables do not need to be renamed in SQLDF. Caution: the function calculating standard deviation is called differently.

Missing grouping

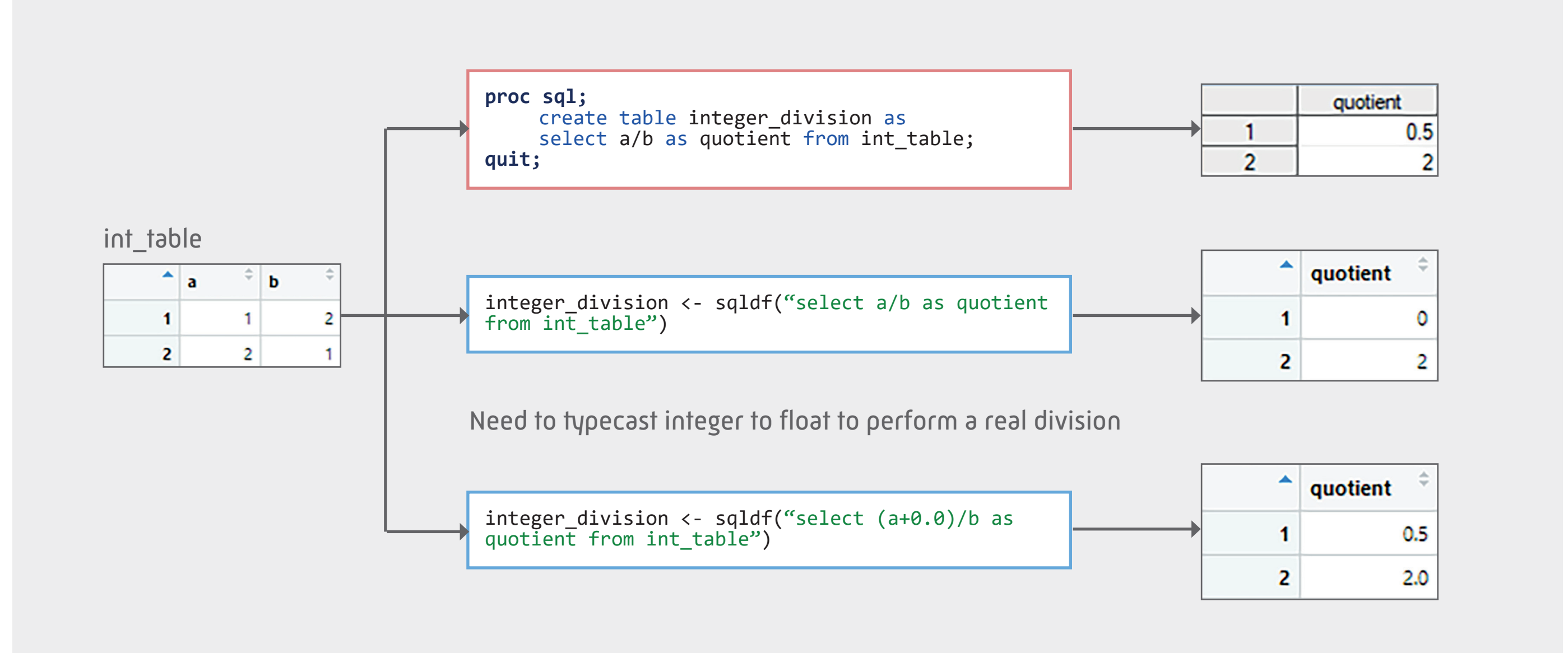


If not all variables from the „select“ statement are added to the „group by“ statement, SQLDF only displays the observations containing the value, calculated by the aggregate function, while PROC SQL displays all observations.

Easy column renaming in R



Integer division leads to integer value in R



Conclusion

- Learning SQLDF with previous knowledge of PROC SQL is not a big effort
- There are some differences to be considered
- Be aware of which database management system SQLDF is using in the backend

Reproduce
R-Code

